Needed: A Science of Open Science for NIH

Michael F. Huerta, Ph.D.
Associate Director for Program Development
National Library of Medicine, NIH

Science of Science Innovation Policy – April 7, 2016
Sharing Data  Open Science
Data are shared in some domains (eg, genomics)

These have an infrastructure for sharing data:
- Clear, heeded policies of funders, publishers, etc.
- Widely-used data-related standards
- Data repositories, platforms & tools

Data infrastructure → Use of FAIR principles for data
- Findable
- Accessible
- Interoperable
- Reusable

Data-centric Open → Progress Success
This cycle is good for science

Sharing Data → Open Science
Most domains of biomedicine are not data-centric nor open.
For these **non-data-centric** domains, the major public products of research are **scientific papers** that describe the authors’ **ideas about** the data....
...but the underlying data are never seen.
This is about to change
Increasing Access to US Research Results

- February 2013 - OSTP Memo to increase public access to results of federal government-supported research
  - Federal agencies > $100M/year research
  - Publications
  - Data

- February 2015 - NIH plan posted
  - Implementation planning is underway
Societal expectations

Technical capabilities

Policy directives

Scientific opportunities

Broad Data Sharing Across All Domains
Broadly Increasing Access to Data

- For **non-data-centric** domains, we have
  - An **imminent flood** of data with
  - Little **infrastructure** to handle the flood
Broadly Increasing Access to Data

Three Facts

- Implementation of NIH Plan will require changes in NIH policies & practices related to data & infrastructure.

- Changes will require consequential decisions.

- Evidence-based decisions are better than those based on anecdote & opinion.

A Science of Open Science can provide that evidence base.
Qs for a Science of Open Science for NIH

- By whom, how, how often and for what purposes are currently available data re-used?
- For specified data repositories, what would be the minimal level of functionality and cost?
- For specified research domains, what infrastructure exists?
- What elements should be considered in a cost/benefit analysis for decisions about infrastructure investment?
- When should infrastructure be owned & operated by NIH versus extramural investigators?
- For specified domains, what is the incentive structure across the research enterprise to encourage specified behaviors (eg, sharing)? How can incentives be systematically aligned?
Science of Open Science for NIH

- Evidence should come from:
  - Administrative systems
  - Studies conducted by appropriate experts
  - Studies directed to address needs of NIH decisions
    - So, actually an applied science of open science

- Evidence base about data & infrastructure
  - **Formative** - Current data & infra landscape
    - Biomedical research
    - Other research areas & sectors
  - **Summative** – Ongoing, to inform course corrections
  - **Facilitative** – Study incentives to promote & shape change from closed to open science
Science of Open Science for NIH

- Identify Key Decisions
- Incentive Strategies
- Empirical Evidence
- Decision Strategies
- Decision Support
Science of Open Science for NIH

Identify Key Decisions

Anecdote Opinion

Incentive Strategies

Decision Strategies

Decision Support